PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07K 14/505, A61K 38/18 | A1

(11) International Publication Number: WO 97/49729

(43) International Publication Date: 31 December 1997 (31.12.97)

(21) International Application Number: PCT/EP97/03228

(22) International Filing Date: 20 June 1997 (20.06.97)

(30) Priority Data:
9613289.9 25 June 1996 (25.06.96) GB

(71) Applicant (*for all designated States except US*): DIAPHARM LIMITED [GB/–]; Quay House, South Esplanade, St. Peter Port, Guernsey GY1 4EJ (GB).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): MARKOVIC, Dejan [YU/IT]; Via Andrea Verga, 5, I-20144 Milano (IT). VELJKOVIC, Veljiko [YU/YU]; Kosovska 15, YU-11000 Beograd (YU). METLAS, Radmila [YU/IT]; Via Colvera, 2, I-33081 Aviano (IT).

(74) Agent: MINOJA, Fabrizio; Studio Consulenza Brevettuale, Via Rossini, 8, I-20122 Milano (IT).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: POLYPEPTIDES MIMICKING THE ACTIVITY OF HUMAN ERYTHROPOIETIN

(57) Abstract

Polypeptides or erythropoietin muteins, characterised by an amino acid sequence which is different from that of human erythropoietin and presents in the informational spectrum obtained by Fourier transformation according to the method of informational analyses, substantially the same frequencies of natural erythropoietin.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

POLYPEPTIDES MIMICKING THE ACTIVITY OF HUMAN ERYTHROPOIETIN

The present invention refers to polypeptides having a sequence obtained by the informational spectra method and the ability of stimulating the production of reticulocytes and red blood cells from bone marrow cells as well as hemoglobin synthesis and iron uptake.

Human erythropoietin (EPO), an hormone playing a fundamental role in erythropoiesis, is a glycoprotein having a molecular weight of about 34-38 kd, the primary structure of which is shown in Figure 1.

EPO is presently obtained by recombinant DNA techniques using eukaryotic cells which can glycosylate the expression product.

The availability of shorter polypeptides, possibly active even in unglycosylated form mimicking the biological activity of human EPO, would be an highly desirable goal allowing the convenient preparation by synthetic procedures.

It has now been found that polypeptides, having in the informational spectrum obtained by Fourier transformation according to the informational analysis method substantially the same frequencies of natural erythropoietin, have substantially the same activity of human erythropoietin.

The informational analysis method (ISM), first disclosed by Veljkovic V. et al. in IEEE Trans. Biomed. Eng. 32, 337 (1985); Cancer Biochem. Biophys. 9, 139, 1987; Phys. Rev. Lett. 29, 105, (1972) and Phys. Lett. 45A, 41, (1973) the content of which is herein

incorporated by reference, is based on the analysis of the information encoded in primary structure which is expressed by molecular electric oscillations propagating through polar environment. Based on the previously demonstrated strong correlation between electron-ion interaction potential (hereinafter EIIP) [Veljkovic V., A theoretical approach to preselection of carcinogens and chemical carcinogenesis, Gordon & Breach Sci. Pub., New York, 1980; Politzer P. and Truhlar D. G., Chemical applications of atomic and molecular electrostatic potential, Plenum Press, New York, 1981); Politzer P., Toxicol, Lett. 43, 227 (1988)] it has been proposed that information expressed by electric oscillations is encoded in protein primary structure by distribution of the values of EIIP of amino acids.

According to this approach, the protein sequences are transformed into signals by assignment of numerical values to each amino acid. These values correspond to EIIP [Veljkovic V. and Slavic I., Phys. Rev. Lett., 29, 105 (1972); Veljkovic V. Phys. Lett., 45A, 41 (1973)]. The signal obtained is than decomposed in periodical function by Fourier transformation. The result is a series of frequencies and their amplitudes. The obtained frequencies correspond to the distribution of structural motifs with defined physico-chemical characteristics responsible for biological function of protein. When comparing proteins which share the same biological or biochemical function, the technique allows detection of code/frequency pairs which are specific for their common biological properties. The method is insensitive to the location of the motifs and, thus, does not require

previous alignment of the sequence. The ISM was successfully applied in structure/function analysis and de novo design of peptides [Cosic I. and Nesis D., Eur.J. Biochem., 170, 247 (1988); Skerl V., and Pavlovic M., FEBS Lett., 239, 1411 (1988); Veljkovic V. and Metlas R., Cancer Biochem. Biophys., 10, 191 (1988); Cosic I. et al., Biochemie, 71, 333 (1989); Lalovic D. and Veljkovic V., Biosystems, 23, 311 (1989); Cosic I., Resonant recognition model of protein-protein and protein-DNA recognition, in Bioinstrumentation and Biosensor (edited by Weis D.L.), Marcel Dekker, Inc., New York (1990); Cosic I. et al., Eur. J. Biochem. 198, 113 (1991); Cosic I. and Hearn M.T. W., J. Mol. Recognition., 4, 57 (1991); Veljkovic V. et al., Biochem. Biophys. Res. Commun., 189, 705 (1992); Krsmanovic B. et al., WO 93/17108.

An object of the invention is provided by EPO muteins having an homology degree with natural erythropoietin lower than 60% or polypeptides having from 20 to 100, preferably from 15 to 70 amino acids, characterized by informational spectrum having substantially the same frequencies as found in the informational spectrum of natural erythropoietin.

The muteins or polypeptides according to the invention may be designed so as to include appropriate O- or N-glycosilation sites even though it has been surprisingly found that glycosilation is not always necessary for the biological activity.

More particularly, the muteins or polypeptides of the invention are characterized by the frequency component 0.312 ± 0.004 in the informational spectrum

and at least one of the following frequency components: 0.023, 0.156, 0.180, 0.185, 0.258, 0.273, 0.285, 0.363 and 0.500 determined with the accuracy of I 0.004.

Preferred muteins of the invention are shown in Figures 2a and 2b (Sequence Id n. 1 and 2) whereas preferred polypeptides are shown in Figure 3 (Sequence Id n. 3, 4, 5 and 6). In Figures 1, 2a and 2b, the predicted amphipathic α helices are underlined with double line (====). Residues predicted to be on the surface are underlined. Each of the three N-linked and one of O-linked glycosylation sites is designated by an asterisk. The mutations are designated by small letters.

Another group of preferred peptides according to the invention and having 16-40 aminoacids, is represented by the following general formula: X1-Tyr-X2-Cys-X3-X4-Gly-Pro-X5-Thr-Trp-X6-Cys-X7-Pro-X8, where X1 = Thr, Gln, His or Asp, X2 = Ser, Asn, Gly, Pro or Ala, X3 = Thr, Gln, Asp, Ser, Asn, Gly, Pro, Ala, Ile, Val, Phe, Lys, Tyr, Met or Glu, X4 = Phe, Val or Met, X5 = Leu, Ile, Met or Val, X6 = Leu, Ile or Val, X7 = Lys, Arg or Asp, and X8 = Gln, Thr, His, Ser, Ala, Val or Leu. Optionally the peptide may be cyclised or dimerised.

These peptides are able to bind to the erythropoietin receptor and show in the IS the frequency component 0.312 and the corresponding amplitude A $\geq$ 0.11.

The invention also refers to polynucleotide sequences coding for said muteins or polypeptides, to expression vectors comprising said polynucleotide sequences and to hosts transformed or transfected by

said vectors, as well as to nucleotide sequences which hybridize to the above mentioned coding sequences.

The polypeptides sequence of the invention are determined by a procedure involving the following steps:

1. determination of the consensus characteristic frequencies for the EPO molecules;
2. derivation of a new numerical sequence of the desired length having the same characteristic frequencies using inverse Fourier transformation;
3. determination of the amino acid corresponding to each element of this new numerical sequence from values of EIIP (see Table 1):

Table 1

Amino acid	EIIP [Ry]*
L	0.0000
I	0.0000
N	0.0036
G	0.0050
V	0.0057
E	0.0058
P	0.0198
H	0.0242
K	0.0371
A	0.0373
Y	0.0516
W	0.0548
Q	0.0761
M	0.0823
S	0.0829
C	0.0829
T	0.0941
F	0.0946
R	0.0959
D	0.1263

* Ry = Rydberg unit

The polypeptides may be obtained by conventional methods of peptide synthesis or by known recombinant DNA techniques.

The polypeptides or muteins of the invention may be administered to humans or animals in form of suitable pharmaceutical compositions, usually but not exclusively

to be administered parenterally. Said compositions will contain from 1 to about 100 mg of mutein or polypeptide for the treatment of the same pathological conditions presently treated with human or recombinant EPO.

The following examples further illustrate the invention.

Example 1

Analysis of amino acid sequences by the ISM

In the first step of the ISM analysis each constitutive element (amino acid) in analyzed sequence is represented by corresponding EIIP value. For calculation of EIIP the following expression derived from the "general model pseudopotential" [Veljkovic V. and Slavic I., Phys. Rev. Lett., 29, 105 (1972); Veljkovic V., Phys. Lett., 45A, 41 (1973)] was used:

$$W = 0.25\ Z^* \sin(1.04\ \pi\ Z^*)/2 \qquad (1)$$

where Z^* is the average quasi-valence number determined by:

$$Z^* = \sum_{i=1}^{m} n_i Z_1 / N \qquad (2)$$

where Z is the number of valence electrons of the i-th atomic component, n_i - the number of atoms of the i-th component, m and N - the number of atomic components and total number of atoms in the side group, respectively. The values of EIIP for side groups of amino acids calculated in accord with Eq. (1) are given in Table 1.

The numerical serie determined in this way is finite-lenght deterministic discrete signal containing information corresponding to selective long-distance interaction among biological macromolecules. In order to analyze this information, the obtained numerical sequence was subjected to discrete Fourier

transformation (DFT), which is defined as follows [Rabiner R., L. and old B., *Theory and applications of digital processing*. Prentice-Hall Inc., nglewood 1975)]:

$$X(n) = \sum_{m=0}^{N-1} x(m)e^{-j\frac{2}{N}nm}, \quad n = 1,2,\ldots\ldots, N/2 \qquad (3)$$

Here x(m) is the m-th of a given numerical series, and X(n) are coefficients of DFT. The coefficients are describing the amplitude, phase, and frequency of sinusoids from which original signal consists. The absolute value of complex DFT coefficients determines the amplitude spectrum which is in the ISM defined as *informational spectrum* (IS) and represented by the following equation:

$$S(n) = X(n)X^*(n) = |X(n)|^2, \quad n^{**} 1, 2,\ldots\ldots,N/2, \qquad (4)$$

It was assumed that points in analyzed numerical sequences are equidistant with the distance d = 1. In this case the maximal frequency in the spectrum is F_{max} = 1/2d = 0.5. It is important to note that the frequency range is independent of number of points in the sequence. The total number of points in the sequence (i.e. number of amino acids in the analyzed primary structure) influences only the resolution of IS. In the case of an N-point sequence, the resolution equals 1/N. The minimal lenght of sequence that can be analyzed by ISM is determined by the desired resolution of the spectrum. Therefore, this number is determined by the expected number of peaks which are to be strictly separated and cannot be exactly defined. The minimal length of sequence which can be analyzed by ISM with suitable accuracy is 16 amino acids.

In this way, the information primarily defined by the sequence of symbols representing amino acids is presented in spectral form which is more suitable for mathematical analysis. It is important to note that ISM does not influence this information and represents only a tool for its analysis (like the prism which decomposes the white light in its spectral components). Each frequency in the IS represents a particular informational component encoded in the primary structure by regularly distributed structural motifs with similar electronic properties.

Example 2

Application of the technique in example 1 to the analysis of EPO proteins

The algorithmic procedures were applied to the mammalian EPO protein sequences.

The analysis procedure comprises the following steps:

1. each amino acid sequences was converted to the numerical sequence by representing each amino acid with the corresponding value of the EIIP;
2. this numerical sequence was converted into a numerical spectrum using fast Fourier transform (hereinafter FFT);
3. spectra were mutually compared using cross-spectral analysis with the aim to extract common frequency components.

From the analysis of cross-spectra of EPO molecules from various mammalian species (mouse, rat, rabbit, ship, monkey and human) one can deduce a set of characteristic frequencies which predominates in the

obtained CIS. In Table 2 all characteristics frequencies in the EPO CIS with S/N > 1 are given:

Table 2

F1	0.023
F2	0.156
F3	0.180
F4	0.195
F5	0.258
F6	0.273
F7	0.285
F8	0.312
F9	0.363
F10	0.500

All frequency components in Table 2 are determined with accuracy of ± 0.004.

From the obtained results, it is possible to conclude that information which is essential for biological activity of the analyzed EPO molecules is completely determined with the set of characteristic frequencies given in Table 2.

Example 3

Application of the technique in example 1 to the analysis of EPO-EPOR interaction

In order to determine which of the characteristic frequency components from Table 2 determines the information that is essential for human EPO-EPOR interaction, the cross-spectrum between these two proteins is obtained.

The characteristic frequencies corresponding to the first 15 amplitudes in EPO-EPOR cross-spectrum are given in Table 3.

Table 3

F	S/N
0.311	10.6
0.258	7.5
0.272	7.3
0.113	7.0
0.361	7.0
0.498	5.2
0.430	5.1
0.158	5.1
0.031	4.6
0.382	4.6
0.154	4.4
0.283	4.3
0.275	3.9
0.347	3.9
0.008	3.5

The results presented in Table 3 show that the main part of information corresponding to human EPO-EPOR interaction is determined by the frequency component 0.311. It is also important to note that, taking into account accuracy of ± 0.004 in the determination of frequency values, 8 of 10 characteristic frequencies from the EPO CIS (Table 2) are also contained within the first 15 characteristic frequencies in the human EPO-EPOR cross-spectrum (Table 3).

Example 4

Application of the technique in example 1 to design of EPO muteins

Once the characteristic frequencies for EPO protein family had been determined, it was possible to design

EPO muteins introducing a large number of amino acid substitutions in human EPO. The main condition that must be satisfied in the design of these muteins is the conservation of IS of human EPO.

In Figure 2a, the primary structure of mutein-1 generated by substitution of 99 (56.6%) amino acids in human EPO is given (Sequence Id n. 1). The IS of this mutein is given in Table 4 and Figure 4a. As can be seen, this IS contains all 10 characteristic EPO frequencies from Table 2, as well as other frequency components corresponding to first 15 amplitudes in IS of the human EPO.

Table 4

IS (EPO)	IS (mut. 1)	IS (mut. 2)
0.359	0.359	0.359
0.500	0.156	0.500
0.156	0.500	0.156
0.195	0.285	0.285
0.258	0.258	0.258
0.285	0.195	0.195
0.203	0.203	0.273
0.273	0.273	0.203
0.180	0.402	0.180
0.004	0.180	0.301
0.445	0.004	0.004
0.113	0.113	0.211
0.023	0.172	0.113
0.312	0.211	0.312
0.050	0.312	0.402

In order to design the mutein that with higher probability will express EPO activity, also structural

characteristics that are important for this biological activity must be taken into account [Boissel JP. et al., J. Biol. Chem., 268, 15983 (1993); Wen D., et al., J. Biol. Chem., 269, 22839 (1994)]. In Figure 2b (Sequence Id n. 2), the primary structure of mutein-2 generated by substitution of 72 (43.4%) amino acids in human EPO is given. This mutein besides conserved IS of human EPO (Table 4 and Figure 4b) also has preserved all structural elements that are important for the biological activity of the human EPO, including all glycosilation sites [Wen D., et al., J. Biol. Chem. 269, 22839 (1994)]. The comparison of the alpha and beta propensity, hydrophilicity, hydrophobicity, solvent accessibility, antigenicity and secondary structure of the human EPO and mutein-2 are given in Figure 5-6.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: DIAPHARM LIMITED
(B) STREET: Quay House, South Esplanade
(C) CITY: St. Peter Port
(D) STATE: Guernsey
(E) COUNTRY: Channel Islands
(F) POSTAL CODE (ZIP): GY1 4EJ

(i) APPLICANT:
(A) NAME: MARKOVIC Dejan
(B) STREET: Via Andrea Verga, 5
(C) CITY: Milano
(D) STATE: Italy
(E) COUNTRY: IT
(F) POSTAL CODE (ZIP): I-20144

(ii) TITLE OF INVENTION: POLYPEPTIDES MIMICKING THE ACTIVITY OF HUMAN ERYTHROPOIETIN

(iii) NUMBER OF SEQUENCES: 6

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version

#1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 166 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Trp Gly Gly Arg Val Val Cys Asp Thr Arg Ile Val Glu Arg
1 5 10

Tyr Val Val Glu Trp His Glu Trp Glu Asn Ile Ser Ser Pro
15 20 25

Cys Trp Glu Lys Cys Thr Val Asn Glu Asn Val Ser Ile Gly
30 35 40

Asp Ser His Ile Asn Phe Tyr Trp Ala His Arg Met Glu Ile
45 50 55

Pro Gln Gln Trp Leu Glu Ile Ala Gln Pro Val Trp Val Val
60 65 70

Thr Glu Trp Ile Val Arg Pro Gln Trp Val Val Val Asn Ser
75 80

Thr Gln Gly Ala Glu Gly Val Gln Val Lys Leu Asp His Trp
85 90 95

Ile Thr Pro Val Arg Thr Val Ser Ser Val Val Arg Trp Val
100 105 110

Pro Trp Gln His Glu Trp Val Thr Gly Gly Asp Ala Ala Ser
115 120 125

Ala Ala Gly Val Arg Ser Val Ser Trp Asp Ser Phe Arg His
130 135 140

Val Phe Arg Leu Tyr Thr Asn Phe Val Arg Gly His Val His
145 150

Val Tyr Ser Gly Glu Trp Cys Arg Ser Pro Asp Arg
155 160 165

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 166 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Trp Gly Gly Arg Val Val Cys Asp Ser Arg Val Leu Glu Arg
1 5 10

Tyr Leu Leu Glu Ala His Glu Ala Glu Asn Ile Thr Ser Pro
15 20 25

Cys Trp Glu Lys Cys Thr Val Asn Glu Asn Ile Thr Ile Pro
30 35 40

Asp Ser His Ile Asn Phe Tyr Trp Ala His Arg Met Glu Ile
45 50 55

Pro Gln Gln Ala Leu Glu Leu Trp Gln Gly Ala Leu Leu Leu
60 65 70

Thr Glu Ala Leu Val Arg Pro Gln Trp Val Leu Val Asn Ser
75 80 85

Ser Gln Gly Ala Glu Gly Leu Gln Leu His Leu Asp His Ala
90 95

Leu Thr Pro Leu Arg Thr Leu Ser Ser Val Val Arg Trp Val
100 105 110

Pro Trp Gln His Glu Trp Val Thr Gly Gly Asp Ala Ala Ser
115 120 125

Ala Ala Gly Val Arg Ser Leu Ser Ala Asp Ser Phe Arg Lys
130 135 140

Val Phe Arg Leu Tyr Thr Asn Phe Leu Arg Gly His Val His
145 150

Val Tyr Ser Gly Glu Trp Cys Arg Ser Gly Asp Arg
155 160 165

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Asp Pro Gln Thr Leu Val Asn Ser Ser Tyr Gln Lys His Asp
1 5 10

Tyr His Asp Pro Leu Asp Ala His Asp His Glu
15 20 25

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Phe His Lys Arg Trp Ala Ala Ser Ala Ala Gln Trp Trp Thr
1 5 10

Ala His Arg Trp Met Phe Gly Tyr Asp Trp Lys Gln His Trp
15 20 25

Asp Glu Asn Ile Asp Gln Ile
30 35

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 64 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Thr Gln Glu Trp Ala Gln Met Ala Ser Ala Phe Ala Trp Arg
1 5 10

Ser His Arg Gln Ala Glu Asn Ile Asp Ala His Thr Glu Gln
15 20 25

Thr Ala His Lys Asp Ser Lys Met Gln Leu Ser Phe Lys Leu
30 35 40

Met Gln Thr Trp His Ala Arg Gln Trp Ala Ala Ser Ala Ala
45 50 55

19

Glu Trp Asp Gln Met Gln Leu Gln
60

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 55 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Gln Tyr Tyr Thr Lys Trp Trp Thr Gln Gln Gln Ala Tyr Asp
1 5 10

Thr Tyr Cys Gln Tyr Gln His Met Thr Val Asn Ser Lys Trp
15 20 25

Arg Gln Leu His Asp Arg His Trp Trp Pro Gln Arg Pro Trp
30 35 40

Tyr Trp Gln Ala His Met Cys Trp Tyr Trp Cys Gln Gln
45 50 55

CLAIMS

1. A polypeptide mimicking the activity of human erythropoietin, having an amino acid sequence which is different from that of human erythropoietin and presents substantially the same frequencies as natural erythropoietin in the information spectrum obtained by Fourier transformation according to the method of informational analysis.

2. A polypeptide according to claim 1, which is an erythropoietin mutein having an homology degree with natural human erythropoietin lower than 60%.

3. A polypeptide according to claim 1 or 2, having from 15 to 70 amino acids.

4. A polypeptide according to any one of claims 1 to 3, having the frequency component 0.312 ± 0,004 in the informational spectrum and at least one of the following frequency components: 0.023, 0.156, 0.180, 0.195, 0.258, 0.273, 0.285, 0.363 and 0.500 the accuracy being ± 0.004.

5. A polypeptide according to any preceding claim, having any of the primary sequences of sequences Id n. 1-6.

6. A polypeptide according to claim 1 including the amino acid sequence: X1-Tyr-X2-Cys-X3-X4-Gly-Pro-X5-Thr-Trp-X6-Cys-X7-Pro-X8, where X1 = Thr, Gln, His or Asp, X2 = Ser, Asn, Gly, Pro or Ala, X3 = Thr, Gln, Asp, Ser, Asn, Gly, Pro, Ala, Ile, Val, Phe, Lys, Tyr, Met or Glu, X4 = Phe, Val or Met, X5 = Leu, Ile, Met or Val, X6 = Leu, Ile or Val, X7 = Lys, Arg or Asp, and X8 = Gln, Thr, His, Ser, Ala, Val or Leu.

7. A polynucleotide coding for a polypeptides according to any preceding claim.

8. An expression vector comprising a polynucleotide according to claim 6.

9. A pharmaceutical composition containing a polypeptide according to any of claims 1 to 5 in admixture with one or more suitable carriers or excipients therefor.

10. A peptide having human EPO activity but which is not natural human EPO, substantially as described herein.

Fig. 1

ERYTHROPOIETIN (human)

```
                10                  20      *       30                  *
APPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNEN
         ==================

40                  50                  60                  70
ITVPDTKVNFYAWKRMEVGQQAVEVWQGLALLSEA
                                  ==========

                80      *           90                  100                 110
VLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLR
==                ==========          ==========

                    120         *   130                 140
ALGAQKEAISPPDAASAAPLRTITADTFRKLFRVYSNF
                                  ==========

150                 160         166
LRGKLKLYTGEACRTGDR
==
```

Fig. 2a

MUTEIN 1 - the huma EPO with 99 (59.6%) amino acids substituted

wggRvvCDtRivERYvvEwhEwENIsspCwEkCtvNENvsigDshiNF
YwahRMEipQQwlEiaQpvwvvtEwivRpQwvvVNStQgaEgvQvkl
DhwitpvRtvssvvRwvpwQhEwvtggDAASAAgvRsvswDsFRhvFR
lYtNFvRGhvhvYsGEwCRspDR

Fig. 2b

MUTEIN 2 - the human EPO with 72 (43.4%) amino acids substituted and conserved α-structure and glycolisation sites

wggRvvCDSRVLERYLLEAhEAENITspCwEkCtvNENITipDshi
NFYwahRMEipQQAlElWQGalLLtEAlvRpQwvLVNSSQgaEgL
QLHlDhAltpLRtLssvvRwvpwQhEwvtggDAASAAgvRslsADsF
RKvFRlYtNFLRGhvhvYsGEwCRsgDR

Fig.3

Peptide 1

DPQTLVNSSYQKHDYHDPLDAHDHE

Peptide 2

FHKRWAASAAQWWTAHRWMFGYDWKQHWDENIDQI

Peptide 3

TQEWAQMASAFAWRSHRQAENIDAHTEQTAHKDSKMQLSFKLMQT
WHARQWAASAAEWDQMQLQ

Peptide 4

QYYTKWWTQQQAYDTYCQYQHMTVNSKWRQLHDRHWWPQRPWY
WQAHMCWYWCQQ

Fig.4a

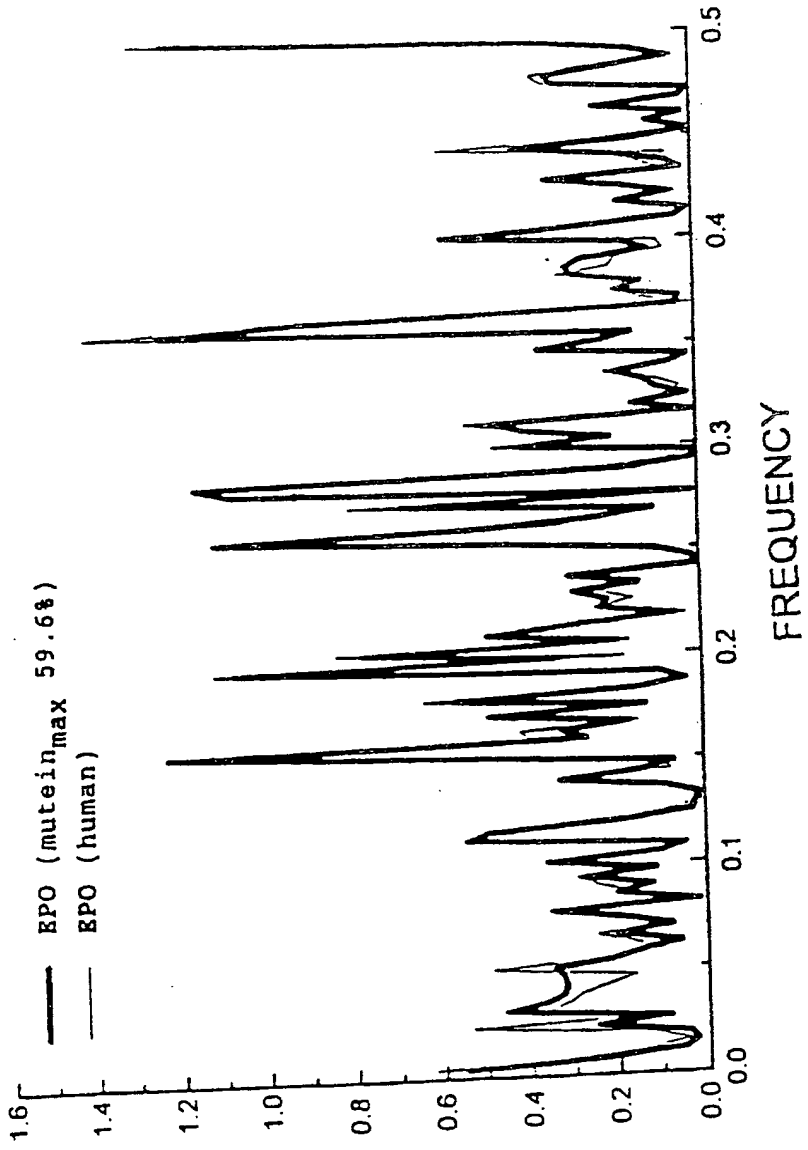

EPO (mutein$_{max}$ 59.6%)
EPO (human)
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.1
0.2
0.3
0.4
0.5
FREQUENCY

Fig.4b

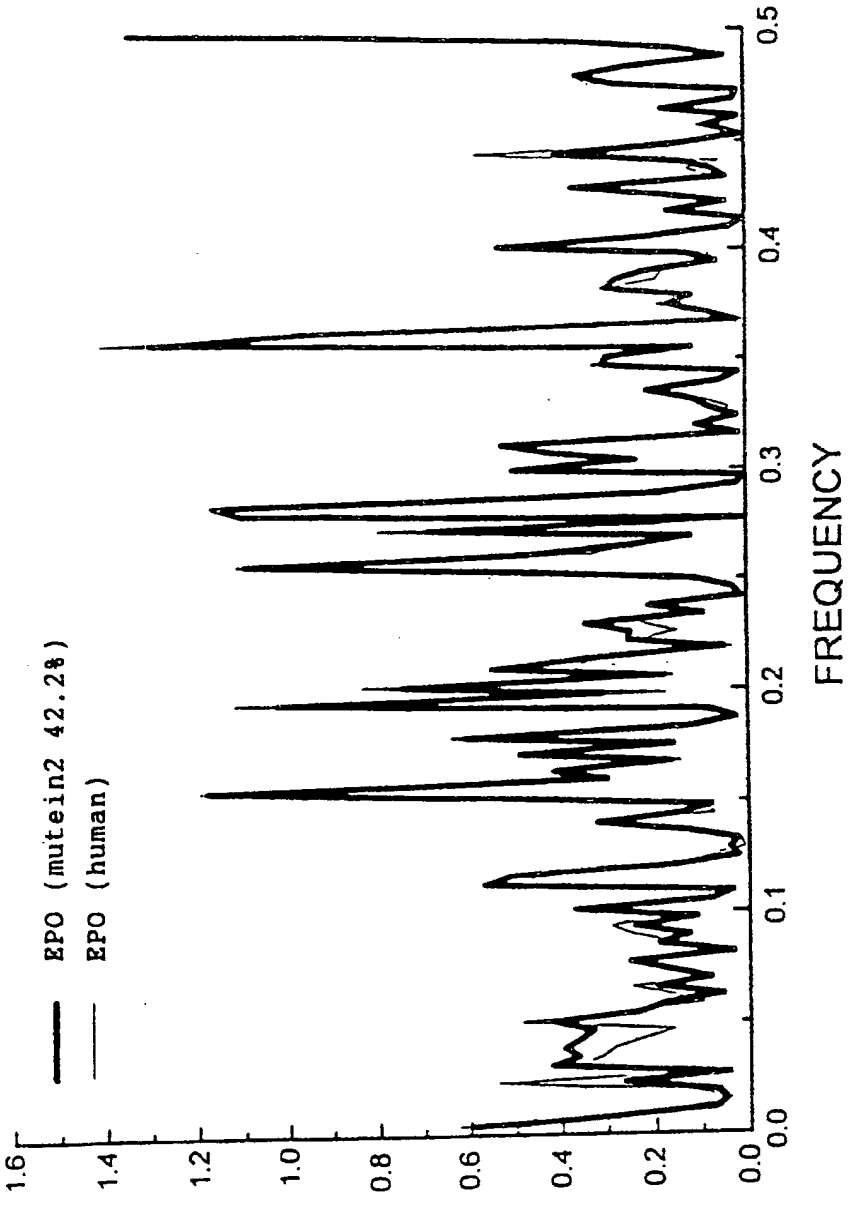

EPO (mutein2 42.2%)
EPO (human)
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.1
0.2
0.3
0.4
0.5
FREQUENCY

Fig.5

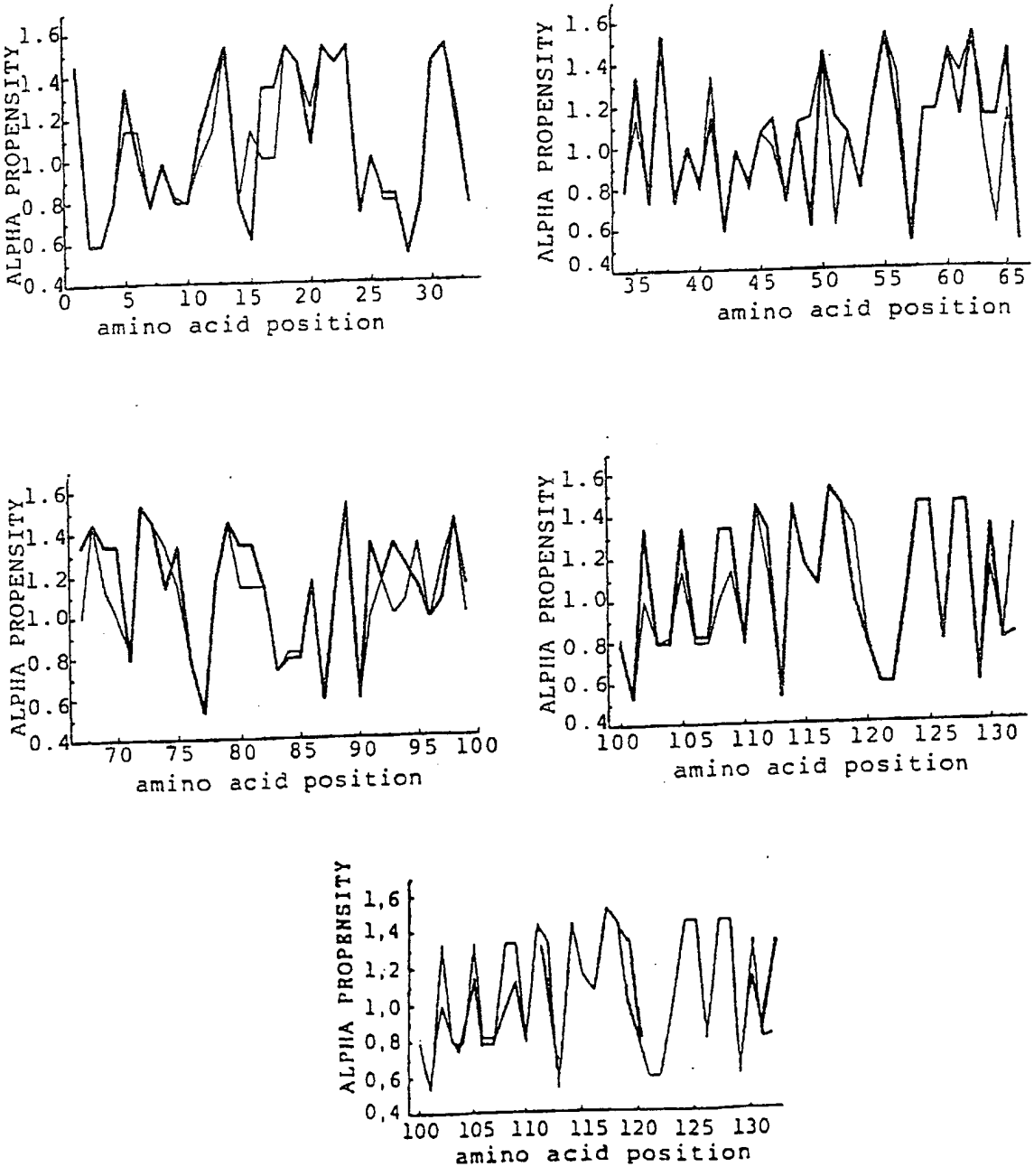
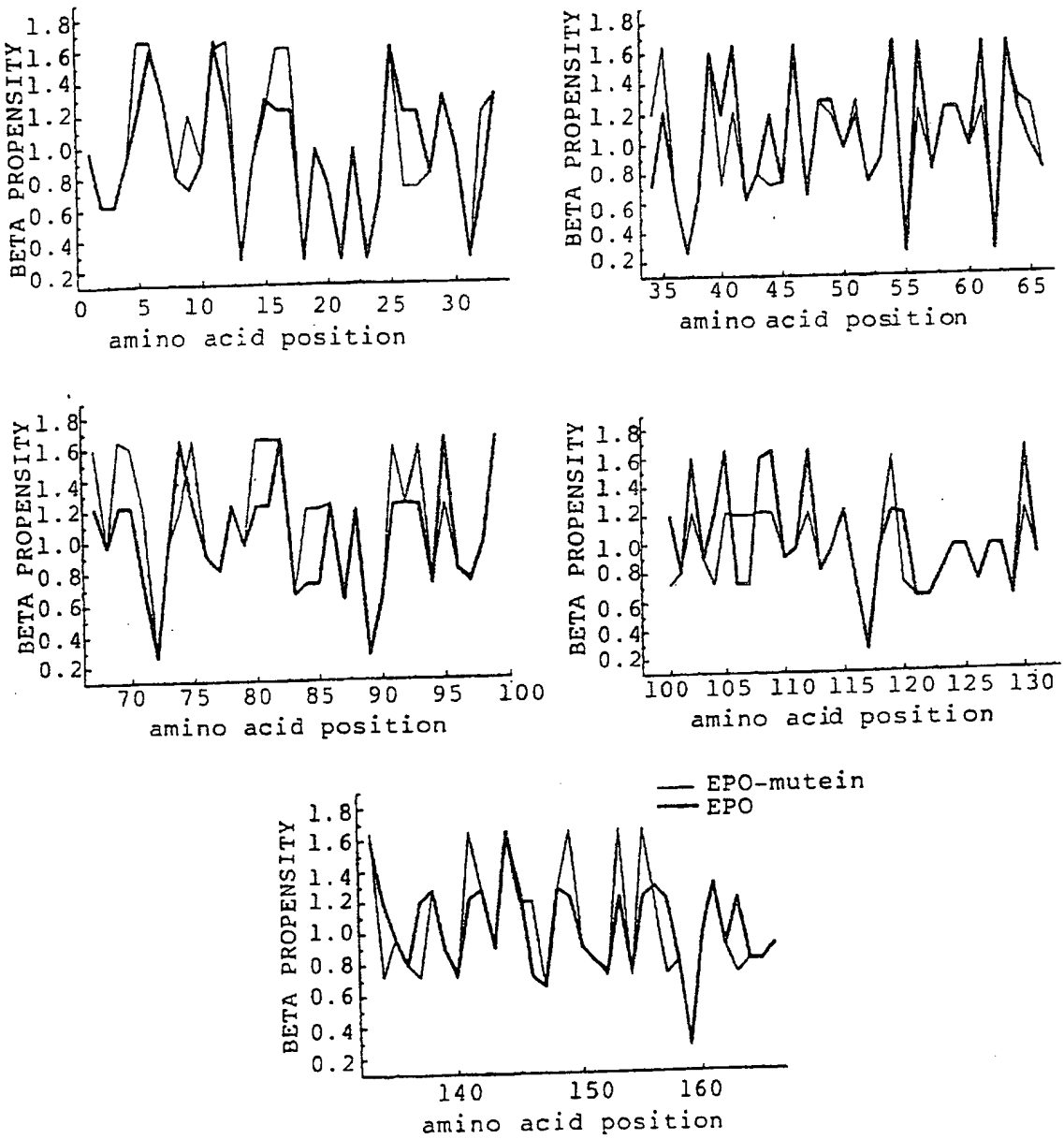

ALPHA PROPENSITY
amino acid position
ALPHA PROPENSITY
amino acid position
ALPHA PROPENSITY
amino acid position
ALPHA PROPENSITY
amino acid position
ALPHA PROPENSITY
amino acid position

Fig.6



INTERNATIONAL SEARCH REPORT

Int ional Application No
PCT/EP 97/03228

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C07K14/505 A61K38/18

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	COSIC I: "MACROMOLECULAR BIOACTIVITY: IS IT RESONANT INTERACTION BETWEEN MACROMOLECULES? - THEORY AND APPLICATIONS" IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, vol. 41, no. 12, 1 December 1994, pages 1101-1114, XP000556764 see the whole document ---	
A	VELJKOVIC ET AL.: "It is possible to analise DNA and protein sequences by the method of digital signal processing ?" IEEE TRANSACTION ON BIOMEDICAL ENGINEERING, vol. 32, no. 5, 1985, pages 337-341, XP002042233 see the whole document ---	

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 September 1997	29. 10. 97
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Cervigni, S

2

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/EP 97/03228

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 93 17108 A (CENTRE NAT RECH SCIENT ;UNIV MONASH (AU)) 2 September 1993 see the whole document ---	
A	WO 94 24160 A (BRIGHAM & WOMENS HOSPITAL) 27 October 1994 -----	

2

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

ernational application No.
PCT/EP 97/03228

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

see FURTHER INFORMATION sheet PCT/ISA/210

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.